Centromere Attribute Integration based Chromosome Polarity Assignment

Ronald J. Stanley, M.S.[1,2], James Keller, Ph.D.[1,2],
Charles W. Caldwell, M.D.[3], Ph.D., Paul Gader, Ph.D.[2]
Medical Informatics Group[1], Departments of Computer Engineering and Computer Science[2],
Pathology and Anatomical Science[3]
University of Missouri, Columbia, MO

Automated karyotyping involves evaluating quantified chromosome attributes for proper classification. Chromosome attributes derived from the banding pattern require the correct chromosome polarity for correct banding sequence interpretation. Chromosome polarity is defined in terms of determining the short and long arms of the chromosome using the centromere as the reference point for measuring the chromosome length on both sides of the centromere. In addition to banding sequence interpretation, polarity is used in the chromosome orientation for chromosome repositioning from the metaphase spread to the karyotype. Automated polarity determination is often not performed for classifying chromosomes in the metaphase spread image. Polarity may be determined user interactively, by the system, or not at all. In order to reduce the computational complexity of evaluating banding sequence features using both chromosome ends as reference points, there is a need to improve chromosome polarity determination in automated karyotyping. A centromere attribute integration approach has been developed at the University of Missouri-Columbia which performs correct chromosome polarity assessment at a rate comparable to other studies of 96.1% on a diversified data set.

INTRODUCTION

Automated and nonautomated approaches to chromosome classification utilize a variety of features or attributes. Attributes commonly measured in existing automated and/or semiautomated karyotyping systems include length, centromeric index, and banding pattern descriptors. Many of those features depend on procuring the chromosome's polarity. Chromosome polarity refers to the determination of a chromosome's short and long arms using the centromere as the reference point for measuring the chromosome length on both sides of the centromere. Polarity is utilized for interpreting banding sequence-based features, reducing the computational complexity of automated karyotyping feature evaluation, and repositioning chromosomes from the metaphase spread to the karyotype. Automated polarity determination is often not performed for classifying chromosomes in the metaphase spread image. Polarity may be determined user interactively, by the system, or not at all. In order to reduce the computational complexity of evaluating banding sequence features using both chromosome ends as reference points, there is a need to improve chromosome polarity determination in automated karyotyping.

Approaches used to determine chromosome polarity are often based on locating the centromere. Identifying the centromere uncovers several important features, including the centromeric index, chromosome orientation, and the banding pattern sequence. The ability to improve centromere identification will enhance the potential for chromosome classification.

Commonly used procedures for polarity assignment using the centromere include: 1) user interaction for centromere determination [1-4], 2) using the system computed centromere location for polarity assignment [5]. Polarity has also been found using the first and second moments of inertia [6].

Research leading to chromosome polarity has focused on a centromere attribute integration approach for centromere identification which integrates some of the common used automated techniques for centromere detection [7]. For G-banded isolated chromosomes or metaphase chromosomes, there are many automated approaches used to detect the centromere. Some of these approaches include: 1) analyzing paired concavities along the chromosome contour, 2) finding shape profile minima, and 3) determining width profile minima [1,5,6,8]. Identifying the true centromere position is based on evaluating the shape and width profile minima in conjunction with the relative changes of the shape and width profiles in neighborhoods of candidate centromere positions. Applying the centromere attribute integration approach to the isolated chromosome image library at the University of Missouri-Columbia has produced correct centromere identification results comparable to other studies [1,5].

The centromere identification procedure has been extended to chromosome polarity determination. A procedure has been developed to determine cases where polarity assignment cannot be performed with certainty, resulting in indeterminate cases. Because correct polarity

0195—4210/96/$5.00

assessment reduces automated karyotyping system computational complexity, finding cases where polarity cannot be found with high confidence avoids unnecessary feature evaluation errors. Applying the polarity determination algorithm to the isolated chromosome library at the University of Missouri-Columbia, polarity is correctly determined at a rate of 96.1%. These experimental results compare favorably with other studies [5]. Enhancing the data set for examining centromere attributes should lead to improvements in polarity assignment.

METHOD

Image Type and Acquisition

All of the experiments were performed using G banded chromosomes at 400-550 banding levels. The images were acquired using a Nikon Axioskop (Zeiss, Thornwood, NY), a Cohu charge-coupled device camera (Cohu, San Diego, CA), and a PowerPC computer (Apple Computer, Cupertino, CA) equipped with a PDI capture board (Perceptive Scientific Instruments, League City, TX). All images were stored as TIFF files. The programs for chromosome image segmentation, skeletonization, and feature extraction were written in ANSI C and implemented on Sun Workstations (Sun Microsystems, Mountain View, CA).

Experimental Data Set

For testing the chromosome orientation identification algorithm, 50 G-banded chromosomes per class are used, except for class y where only 28 specimens are available. Isolated chromosome images with their metaphase orientations are used for algorithm and system development and testing.

Feature Extraction Process

Applied to isolated chromosome images, the feature extraction process utilizes three program inputs. The program inputs are: 1) the original grey-level image, 2) the segmented image of the original grey-level image, and 3) the skeletons determined from the segmented image. Figure 1 shows the three images input to the feature extractor.

The segmented images are generated from a three step process. The initial step involves automatically generating and applying a global threshold to the entire input grey-level image, providing an under segmentation of the image. Secondly, for each connected component of the globally thresholded image, an automated local thresholding procedure is used to facilitate object separation. Finally, all image objects are labeled using connected components analysis [9]. Each segmented image serves as the outer boundary for orthogonal line construction from the skeleton. The analyzed skeleton represents the medial axis for the corresponding segmented image. The medial axis approximately bisects the corresponding segmented image, facilitating feature calculations from the grey-level image.

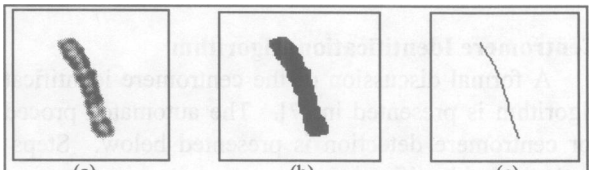

Figure 1: Inputs to the Feature Extractor. (a) Segmented Image. (b) Skeleton Image. (c) Grey-Level Image.

The experimental centromere is identified based on computing the following features: 1) blob width, 2) the mean grey-level along perpendicular lines to the medial axis, and 3) the shape profile along perpendicular lines to the medial axis, and 4) concavity measure based on the shape profile at a given axis point. The blob width is computed as the sum of the Euclidean distances of the axis point to the corresponding perpendicular line end points for each side of the medial axis. The mean is determined as the sum of grey values from the original grey-level image which correspond to the perpendicular line pixels, including the medial axis point grey value, divided by the number of perpendicular line points for the current medial axis point. The shape profile, as determined in other studies [2,3], is the ratio of the sum of each orthogonal line point grey value multiplied by its corresponding squared Euclidean distance from the axis point to the sum of perpendicular line point grey values. The concavity measure provides an indication of the relative change of the width and grey-level in the neighborhood of a given medial axis point. The concavity measure is computed as the sum of the shape profile values of the immediate neighbors of an axis subtracting change of shape profile at the given axis point. The equations implemented for the mean grey-level, shape profile, and concavity measure at each medial axis point over all perpendicular line points are:

$$\text{mean grey-level} = (\sum G(x,y))/n,$$

$$\text{second moment} = (\sum G(x,y)\, d(x,y)^2)/\sum G(x,y),$$

$$\text{concavity measure} = (S(c+2) - 2\,S(c) + S(c-2))/\Delta L, \text{ where}$$

G(x,y): grey-level corresponding to perpendicular line coordinate (x,y),
d(x,y): Euclidean distance medial axis to perpendicular line coordinate (x,y),

n : number of points in perpendicular line to medial axis,

S(c): shape profile at medial axis centromere candidate point c, and

ΔL: change in Euclidean length from sample c+2 to sample c-2.

Centromere Identification Algorithm

A formal discussion of the centromere identification algorithm is presented in [7]. The automated procedure for centromere detection is presented below. Steps for centromere identification:

1. Apply three element median filtering to the density, shape, and width profiles. Only the middle 90% medial axis points are considered for candidate centromeres.
2. Choose initial centromere candidate locations from the shape profile. Candidates are chosen as the three lowest minima which are separated by three or more samples.
3. Find the three lowest minima separated by three or more samples from the width profile.
4. Due to digital constraints associated with chromosome orientation within metaphase spread images, set an upper limit for the width at the square root of two plus the minimum width. Width minima exceeding the upper limit are discarded from centromere candidate assessment. Note that at least the global width minimum remains for further centromere candidate assessment.
5. Cross validate the samples corresponding to the shape profile minima to see if they fall within three samples of the remaining width profile minima. If no shape profile minima fall within three samples of the remaining width profile minima, the centromere is identified as position corresponding to the global shape profile minimum. If only one shape profile minima falls within the width profile range, the centromere is labeled that the location corresponding to that minimum. Otherwise, the positions corresponding to two or three shape profile minima remain as centromere candidates.
6. Compensate the shape profiles of the non-global minima candidates to account for possible rotation distortions, and an upper bound for the shape profile is generated for each remaining candidate. The compensation involves scaling the global shape profile minimum to account for the width variations within the acceptable width profile range. If all candidates exceed the upper bound for the compensated shape profile, the location corresponding to the lowest shape profile minima is designated the centromere. If only one candidate is less than or equal to the upper bound of the compensated shape profile, the centromere is deemed that position. Otherwise two or three candidates remain.
7. Inspect the width profiles of the remaining candidates to see if neighbors have widths below the upper bound for the width previously defined. If a candidate satisfies this condition, the sample corresponding to this candidate is labeled as the centromere. Otherwise, the candidate with the greater local curvature of the shape profile is identified as the centromere. The curvature is defined as the discrete second derivative approximation for the 1-D shape profile.

Chromosome Polarity Assignment Algorithm

Chromosome orientation is determined as identifying the chromosome end point reference corresponding to the short arm. Chromosome orientation is found as follows:

1) Identify the centromere using the centromere identification algorithm previously described.
2) Compute the length of the two chromosome arms in terms of medial axis length between each end point and the centromere position. Length is the sum of Euclidean length increments along the medial axis from the centromere to the respective chromosome end point.
3) Choose the chromosome arm with the shorter length and its end point as the chromosome top. Picking the shorter arm for polarity purposes must satisfy the following arms ratio constraint. Centromeric index (CI) is defined for those studies as the ratio of the short arm length to the total chromosome length. Examining centromeric indice data from [10,11] reveals cases where the CI approaches or exceeds 0.5. Consequently, short and long arm designation for those classes may be indeterminate for correctly identified centromeres. Using the mean and standard deviation centromeric indices found from [10,11], arms ratios which are greater than 0.483 are deemed indeterminate.

Experimental Procedure

The polarity finder was tested using 50 G-banded chromosomes from each class, except for class y where only 28 chromosomes are available. A white dot on the original grey-level image represents the program identified centromere. Polarity assignment makes chromosome top as the end point of the short arm chosen by the program. For situations where the centromere chosen resulted in arms ratios greater than 0.483 or arm lengths, the polarity is indeterminate. A trained cytogenetic technician scored the polarity identification based on the criteria previously described.

RESULTS

In order to illustrate the polarity labeling procedure, Figure 2 presents four sample cases with correctly identified polarities. The single ended black arrow points to the program chosen centromere, designated with a white dot. The double ended light grey arrow represents the chromosome's program chosen short arm. The double ended darker grey arrow represents the

chromosome's chosen long arm. Figure 2 (a) is a chromosome 1 with appropriately identified centromere and polarity. Figure 2 (b) is a chromosome 14 with properly labeled centromere and polarity. Figure 2 (c) is a chromosome 8 with correctly determined centromere and polarity. Figure 2 (d) is a chromosome 9 with incorrectly identified centromere but with properly identified polarity.

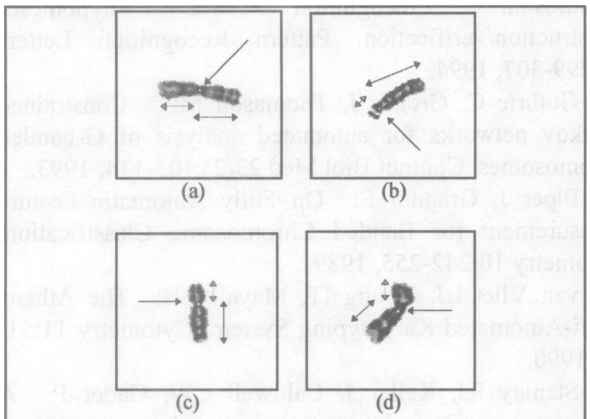

Figure 2: Correctly Identified Chromosome Polarities. (a) Chromosome 1. (b) Chromosome 14. (c) Chromosome 8. (d) Chromosome 9.

Table 1 shows the experimental chromosome polarity identification results. The number of chromosomes tested for each class is presented with the number and percentage of specimens correctly identified. Totals are also presented for each parameter. The results obtained from applying the arms ratio constraint are presented in parentheses.

DISCUSSION

The polarity assignment algorithm is designed to examine chromosomes in their metaphase orientations. The algorithm uses the program chosen centromere as a reference point for finding a chromosome's long and short arms. The correct experimental polarity assignment, 96.1%, compares favorably to other studies [5]. Recognizing indeterminate polarity assignment cases, applying the arms ratio constraint, improves the correct assignment rate. Polarity determination was performed on the isolated chromosome data set at the University of Missouri-Columbia. The algorithm has not been tested on other data sets. There is difficulty assessing the algorithm's performance on other data sets due to differences in preparation techniques, image acquisition procedures, and image enhancement procedures among cytogenetic laboratories. However, the algorithm is directly extendible to highlighting certain chromosome centromere features. The intent of this study is to show that automated polarity assessment is directly related to the ability to automatically find the centromere correctly.

Table 1: Correctly Determined Chromosome Orientation by Class and Total.

Class	Tested	Correct	% Correct
1	50 (42)	44 (41)	88.0 (97.6)
2	50 (50)	47 (47)	94.0 (94.0)
3	50 (43)	47 (41)	94.0 (95.3)
4	50 (50)	49 (49)	98.0 (98.0)
5	50 (50)	49 (49)	98.0 (98.0)
6	50 (50)	47 (47)	94.0 (94.0)
7	50 (50)	49 (49)	98.0 (98.0)
8	50 (50)	46 (46)	92.0 (92.0)
9	50 (50)	48 (50)	96.0 (96.0)
10	50 (50)	47 (47)	94.0 (94.0)
11	50 (50)	49 (49)	98.0 (98.0)
12	50 (50)	48 (48)	96.0 (96.0)
13	50 (50)	48 (48)	96.0 (96.0)
14	50 (50)	49 (49)	98.0 (98.0)
15	50 (50)	47 (47)	94.0 (94.0)
16	50 (50)	48 (48)	96.0 (96.0)
17	50 (50)	46 (46)	92.0 (92.0)
18	50 (49)	48 (48)	96.0 (98.0)
19	50 (45)	46 (43)	92.0 (95.6)
20	50 (46)	49 (45)	98.0 (97.8)
21	50 (50)	49 (49)	98.0 (98.0)
22	50 (50)	50 (50)	100.0 (100.0)
X	50 (48)	46 (45)	92.0 (93.8)
Y	28 (28)	26 (26)	92.9 (92.9)
Total	1178 (1151)	1123 (1106)	95.3 (96.1)

Error sources associated with polarity determination can be characterized as follows. First, incorrect centromere identification not close to the actual centromere position provide the primary error source with proper polarity assessment. As prior research indicates [7], centromere identification errors can be categorized into two cases. First, some of the chromosome specimens are long and narrow without any region characteristic of the centromere. Second, centromere identification errors often occur for chromosomes with centromeres located in regions of extreme bending. For those error sources, centromere attributes are difficult to quantify and to interpret.

In addition to incorrect automated centromere identification, some chromosomes in their metaphase spread positions yield incorrect polarity assessment using the centromere as a reference. The arms ratio constraint, 0.483, is obtained from standard deviation data for chromosomes having CI near 0.5 [10,11]. The standard deviation value chosen is from class 1, the chromosome class which statistically has the highest CI and, thereby, the greatest opportunity for confusing the short and long

arms. Figure 3 is an illustration of a chromosome 1 with the centromere correctly identified and the associated polarity is improperly assessed. The single ended black arrow points to the program chosen centromere, designated with a white dot. The double ended light grey arrow represents the chromosome's program chosen short arm. The double ended darker grey arrow represents the chromosome's chosen long arm.

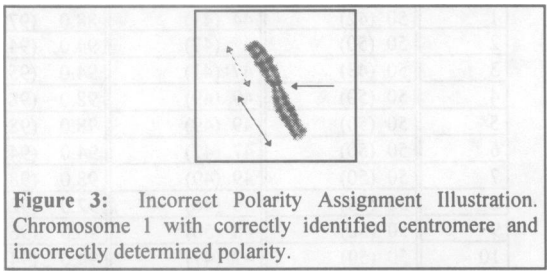

Figure 3: Incorrect Polarity Assignment Illustration. Chromosome 1 with correctly identified centromere and incorrectly determined polarity.

The experimental results provide the basis for two observations. First, the correct polarity assignment rate, 96.1%, from this study is a high classification rate. Highly confident polarity assessment will simplify automated karyotyping systems and will assist in high chromosome classification rates when combined with other features. Secondly, automated polarity determination reduces user interaction with karyotyping systems. The experimental results show that polarity assignment can be performed with high success without user interaction.

Regardless of the approach for classifying chromosomes, there is a need for determining chromosome polarity. Proper chromosome polarity is required for positioning chromosomes from the metaphase spread image to the corresponding karyotype. Additionally, the computational complexity of systems performing automated karyotyping is also dependent on the ability to accurately assess chromosome polarity for chromosomes in the metaphase spread image. The better the ability to determine chromosome polarity the fewer comparisons that are required between features (dependent on polarity) of chromosomes within the metaphase spread. Consequently, the errors in polarity assessment provide the impetus for further research to examine centromere identification and other potential approaches for determining polarity.

Acknowledgments

This research was supported in part by a grant from the University of Missouri Research Board and training grant 5 T15 LM/CA07089-04 from the National Library of Medicine and National Cancer Institute, Bethesda, MD.

References

[1] Graham J: Automation of routine clinical chromosome analysis. Anal Quant Cytol Histol 9:383-390, 1987.

[2] Lundsteen C, Gerdes T, Maahr J: Automated multiple-cell karyotyping: a clinical feasibility study. Clin Genet 39:338-346, 1991.

[3] Denisov DA, Dudkin AK: Model-based chromosome recognition via hypotheses construction/verification. Pattern Recognition Letters 15:299-307, 1994.

[4] Guthrie C, Gregor J, Thomason MG: Constrained Markov networks for automated analysis of G-banded chromosomes. Comput Biol Med 23(2):105-114, 1993.

[5] Piper J, Granum E: On Fully Automatic Feature Measurement for Banded Chromosome Classification. Cytometry 10:242-255, 1989.

[6] van Vliet LJ, Young IT, Mayall BH: The Athena Semi-Automated Karyotyping System. Cytometry 11:51-58, 1990.

[7] Stanley RJ, Keller J, Caldwell CW, Gader P: A centromere attribute integration approach to centromere identification. Publication pending conference proceedings Rocky Mountain Bioengineering Symposium April 1996.

[8] Piper J: Finding chromosome centromeres using boundary and density information. In: *Digital Image Processing*, Simon J-C, Haralick RM (eds.). D Reidel, Dordrecht, The Netherlands, 1981, pp. 511-518.

[9] Haralick RM, Shapiro LG. *Computer and Robot Vision*, New York: Addison-Wesley, 1992.

[10] Mark HF, Mark R, Pan T, Mark Y: Centromere index derivation by a novel and convenient approach. Ann Clin Lab Sci 23(4):267-274.

[11] ISCN: An International System for Human Cytogenetic Nomenclature: Report of the Standing Committee on Human Cytogenetic Nomenclature (1985). Harnden DG, Klinger HP, Jensen JT, Kaelbling M, eds. Basel, Karger S. Published in collaboration with the March of Dimes Birth Defects Foundation and Cytogenetic Cell Genetics: Appendix 2, 1985.